

**Zeitschrift:** Technische Mitteilungen / Schweizerische Post-, Telefon- und Telegrafienbetriebe = Bulletin technique / Entreprise des postes, téléphones et télégraphes suisses = Bollettino tecnico / Azienda delle poste, dei telefoni e dei telegrafi svizzeri

**Herausgeber:** Schweizerische Post-, Telefon- und Telegrafienbetriebe

**Band:** 73 (1995)

**Heft:** 5

**Artikel:** Service téléphonique automatique de renseignements

**Autor:** Pfister, Beat / Schaub, Arthur

**DOI:** <https://doi.org/10.5169/seals-875939>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 02.05.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# Service téléphonique automatique de renseignements\*

Beat PFISTER, Zurich, et Arthur SCHAUB, Hombrechtikon

## Zusammenfassung

*Automatische Rufnummern-Auskunft*

Sprachgesteuerte automatische Auskunftssysteme kommunizieren mit dem Benutzer über das Medium der gesprochenen Sprache. Das Ziel dieser Kommunikation ist, den Informationswunsch des Benutzers zu eruieren und anschliessend die gewünschten Auskünfte zu geben. Ein solches System braucht nebst den Fähigkeiten, gesprochene Sprache zu verstehen (Spracherkennung) und Sprache zu produzieren (Sprachsynthese) auch Zugriff auf die verlangten Informationen. Dies können beispielsweise Wetterprognosen, Telefonnummern, Aktienkurse, Fahrplaninformationen usw. sein.

## Résumé

*Service téléphonique automatique de renseignements*

Les serveurs vocaux interactifs communiquent oralement avec l'utilisateur. Le but de cette communication est de découvrir quelle information l'utilisateur désire et de lui fournir ensuite ces renseignements. Un tel système doit non seulement comprendre le langage parlé (reconnaissance de la parole) et produire des signaux vocaux (synthèse de la parole), mais également pouvoir accéder aux informations désirées. Il peut s'agir par exemple de la météo, de numéros téléphoniques, des cours de la Bourse, d'informations sur les horaires de transport, etc.

## Riassunto

*Servizio telefonico automatico d'informazione*

I sistemi d'informazione automatici con guida fonica comunicano con l'utente mediante la voce. Lo scopo della comunicazione è di scoprire che cosa desidera il cliente e di dargli quindi le informazioni richieste. Oltre ad essere in grado di capire la voce (riconoscimento della voce) e di generarla (sintesi della voce), un simile sistema deve poter anche accedere alle informazioni richieste. Quest'ultime possono essere di tipo diverso: previsioni del tempo, numeri telefonici, quotazioni di borsa, orari di partenza e di arrivo ecc.

## Summary

*Automatic Telephone Directory Assistance*

Speech controlled automatic information systems communicate with the user by means of the spoken language. The aim of this communication is to find out what the user wants to know and then to provide the required information. Such a system needs, apart from the capabilities of understanding the spoken language (speech recognition) and producing speech (speech synthesis), access to the required information. This could be for example weather forecasts, telephone numbers, stock quotations, timetable information etc.

## Introduction

Les serveurs vocaux interactifs communiquent oralement avec l'utilisateur. Le but de cette communication est de découvrir quelle information l'utilisateur désire et de lui fournir ensuite ces renseignements.

Un tel système doit non seulement comprendre le langage parlé (reconnaissance de la parole) et produire des signaux vocaux (synthèse de la parole), mais également pouvoir accéder aux informations désirées. Il peut s'agir, par exemple, de la météo, de numéros téléphoniques, des cours de la bourse, d'informations sur les horaires de transport, etc.

Il est clair qu'un tel système intéresse particulièrement les Télécom PTT, puisqu'il permet de communiquer automatiquement aux abonnés les numéros téléphoniques désirés. C'est pourquoi, dès 1991, un

projet a été mis en chantier avec la collaboration de l'Ecole polytechnique fédérale de Zurich (EPFZ) et d'Ascom Tech SA. L'objectif consiste à réaliser une première approche vers un système automatique de renseignements fournissant les numéros téléphoniques demandés, en allemand «Automatische Rufnummern-Auskunft» (ARA).

## Le service de renseignements automatique idéal

Dans un cas idéal, le système ARA devrait pouvoir remplacer entièrement un service de renseignements desservi. Cela signifie avant tout que chaque utilisateur devrait être en mesure, sans connaissances préalables, de demander oralement au système l'information désirée. Avec l'accès à une banque de données adéquate, le système chercherait l'information souhaitée et finalement la communiquerait oralement à l'utilisateur.

\* Projet auquel collaborent l'Ecole polytechnique fédérale de Zurich, Ascom Tech SA et les Télécom PTT

On souhaite, en outre, un système automatique fonctionnant rapidement, jamais surchargé, ne faisant aucune erreur et disposant toujours des informations les plus récentes. Il doit être en plus peu coûteux pour l'utilisateur et rentable pour l'exploitant.

Un système de renseignements automatisé, capable de répondre à toutes ces exigences, relève aujourd'hui encore de l'utopie. En effet, il ne saurait être réalisé sans que soient résolus au préalable de nombreux problèmes. Plusieurs d'entre eux proviennent du fait que le langage naturel, à l'opposé d'un langage formel (par ex. un langage de programmation), est extrêmement difficile à traiter:

- La recherche dans le domaine de la reconnaissance de la parole est encore loin d'avoir atteint le niveau nécessaire. Un système avec des performances de reconnaissance comparables à celle de l'homme, c'est-à-dire capable de convertir une phrase parlée en un texte avec un vocabulaire illimité, une syntaxe quelconque et voire d'un dialecte n'est pas réalisable dans un proche avenir.
- Il n'est, par ailleurs, pas suffisant que le système de renseignements convertisse une question prononcée par l'utilisateur en un texte correspondant. Le système doit comprendre ce que l'utilisateur veut et doit savoir comment, dans un cas concret, l'information désirée peut être extraite des banques de données disponibles. La possibilité d'analyser le sens de l'expression orale se limite aujourd'hui encore à des applications simples, c'est-à-dire avec un vocabulaire très restreint et des formulations rudimentaires.
- Chaque communication orale crée des problèmes de compréhension à différents niveaux: un message peut ne pas être écouté, entendu ou compris. Pour qu'une communication réussisse, il faut que la machine reconnaisse ces situations et qu'elle réagisse correctement. D'autre part, l'utilisateur doit pouvoir dire en tout temps à la machine qu'il n'a pas compris ce qu'elle a voulu lui communiquer.
- En plus de ces difficultés inhérentes au langage humain, il s'agit encore de résoudre des problèmes impliquant la connaissance générale. Pour pouvoir donner certains renseignements, il est par exemple nécessaire que la machine reconnaisse les notions relevant du temps et de l'espace (géographie), c'est-à-dire qu'elle soit capable d'interpréter les notions d'antériorité, de postériorité ou de proximité, etc.

Pour qu'un système de renseignements automatique puisse intégralement remplacer une place de travail desservie, il est indispensable que ces problèmes soient résolus. Il ne serait cependant ni nécessaire ni sensé de recourir ici au principe du «tout ou rien». Des systèmes de renseignements automatiques simples sont parfaitement réalisables et fonctionnent déjà à l'étranger. Ils exigent cependant que l'utilisateur s'adapte dans une très large mesure aux possibilités de la machine. Pour l'essentiel, ceci détermine la manière de s'adresser oralement à la machine ou, en d'autres termes, c'est le système de reconnaissance de la parole qui fixe les possibilités d'interaction orale.

Par rapport à la représentation idéale formulée plus haut, les performances des systèmes de reconnaissance de la parole actuellement disponibles sont relativement modestes. On peut considérer aujourd'hui que des systèmes capables de reconnaître des mots isolés, accessibles par l'ensemble des locuteurs, ne portant que sur un vocabulaire comprenant d'un à plusieurs centaines de termes, sont aujourd'hui exploitables. Cependant, un système capable de comprendre des phrases complètes, comme ce serait le cas pour une machine idéale, n'en est encore qu'au stade des essais en laboratoire et leur succès se limite à un vocabulaire très restreint.

Dans ce qui suit, on explique comment une première étape est franchie pour satisfaire les exigences d'un système de renseignements automatiques, tout en s'accommodant des limitations définies dans le cadre du projet ARA.

### *Le but visé par le projet ARA*

Le système ARA permet d'obtenir le numéro d'appel d'un abonné par l'introduction de son nom et de son adresse. A l'inverse, la machine peut aussi communiquer le nom et l'adresse d'un abonné après l'énoncé de son numéro. Avec l'annuaire électronique d'abonnés (ETV) des Télécom PTT, cette consultation est relativement simple à effectuer par le biais d'un terminal de visualisation. Or, pour de nombreux usagers, il serait cependant plus utile de pouvoir questionner l'annuaire verbalement au moyen d'un téléphone normal plutôt que par le truchement d'un clavier et d'un écran.

La raison pour laquelle cette opération n'est pas si simple à réaliser réside notamment dans le fait que le vocabulaire utilisé lors de cette communication n'est pas limité, en raison des données à énoncer (nom et prénom de l'abonné, désignation de la rue et de la localité). On ne peut de ce fait recourir à un dispositif conventionnel de reconnaissance des mots, grâce auquel le système serait en mesure d'interpréter les données d'un abonné dont on désire connaître le numéro.

Pour la première version d'un système ARA (appelé ARA-V1), qui fonctionne avec un dispositif de reconnaissance des mots indépendants du locuteur, on est parti de l'hypothèse que l'utilisateur ne prononce pas les noms propres en continu, mais selon la table d'épellation officielle figurant dans chaque annuaire téléphonique. Il a été ainsi possible de réduire de manière drastique le vocabulaire dont le système ARA a besoin et de se limiter aux mots de la table d'épellation, aux chiffres et à quelques mots de commande ou de dialogue.

Cela étant, il est clair que le système est loin de représenter pour un usager une solution capable de remplacer entièrement une position d'opératrice desservie. Malgré cela, ce premier pas est nécessaire. Au moyen de cette approche simple, il a été possible d'étudier tout aussi bien qu'avec un système beaucoup plus complexe de nombreuses questions relatives à la communication homme-machine ainsi qu'à

l'interconnexion avec l'annuaire électronique d'utilisateurs. Mais un système ARA équipé d'un dispositif de reconnaissance de la parole beaucoup plus évolué devra aussi recourir dans tous les cas à l'épellation intégrale ou partielle de certains noms propres, que ce soit à cause de l'orthographe différente des noms à prononciation identique (par ex. Maier, Meier, Meyer), ou parce que l'intelligibilité est mauvaise.

La figure 1 montre un schéma de principe simplifié d'ARA-V1. Les quatre éléments principaux sont:

#### Le dispositif de reconnaissance des mots

Une unité permettant de reconnaître des mots prononcés par des locuteurs quelconques (voir chap. «Le dispositif de reconnaissance des mots»).

#### Le synthétiseur de la parole

Le synthétiseur de la parole utilisé ici est le logiciel SVOX développé par l'EPF de Zurich. Celui-ci donne des indications à l'utilisateur du système ARA, lui pose des questions ou répond à ses demandes (voir chap. «Le synthétiseur de la parole SVOX»).

#### Interface ETV

Il s'agit d'une interface intelligente avec l'annuaire électronique des abonnés des Télécom PTT (voir chap. «L'interface ETV»).

#### Commande centralisée

Ce bloc définit un processus souple de dialogue et de décision, qui, en recourant à tout le savoir disponible, établit avec l'utilisateur un dialogue ciblé, à l'aide de la reconnaissance de la parole et de la synthèse de la parole.

Grâce à cette configuration de système, le dialogue avec l'utilisateur peut être librement structuré, compte tenu bien sûr des conditions marginales voulant que premièrement le dispositif de reconnaissance de la parole ne puisse reconnaître que certains mots et que deuxièmement ceux-ci soient sujets à des erreurs de reconnaissance plus ou moins fréquentes suivant le locuteur. Dans la conception du dialogue d'ARA-V1, on a de ce fait surtout considéré les aspects suivants:

- Le système ARA prend dès le début l'initiative du dialogue, pour empêcher que l'utilisateur commence à parler librement au lieu d'utiliser les mots de dialogue prévus. De plus, les questions et les indications doivent être formulées de manière que l'utilisateur ne soit pas tenté de répondre par une phrase au lieu d'un mot.
- Le système doit pouvoir être utilisable sans information préalable. C'est ainsi que le nouvel utilisateur est guidé par des indications qui ne devraient pas ennuyer un utilisateur exercé.
- Le dialogue normal doit être aussi simple que possible pour l'utilisateur, se dérouler expéditivement, sans toutefois l'obliger à se hâter.
- Le but du dialogue est d'extraire de la banque de données les inscriptions de l'abonné recherché au moyen d'un nombre minimal d'étapes.
- Le dialogue doit recourir à la procédure des quit-

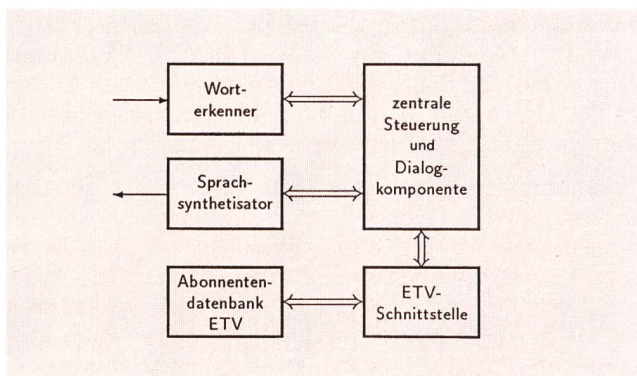


Fig. 1 Schéma bloc du système ARA  
 ETV, annuaire électronique des abonnés  
 Worterkenner – Système de reconnaissance des mots  
 Sprachsynthesator – Synthétiseur de parole  
 Abonentendatenbank ETV – Annuaire électronique des abonnés  
 Zentrale Steuerung und Dialogkomponente – Commande centralisée  
 ETV Schnittstelle – Interface ETV

tances, pour empêcher que le dialogue échoue en cas d'erreurs de reconnaissance de mot. Le principe des quittances et la correction doivent toutefois être prévus de manière que l'utilisateur ne perde pas la vue d'ensemble lorsqu'il épelle des mots ou introduit une série de chiffres.

Quelques-uns de ces aspects se sont révélés antagoniques lors du développement (par exemple, chaque quittance prolonge le dialogue). Dès lors, on ne peut en tenir compte que de manière limitée, et il faut toujours évaluer la priorité de tel ou tel critère à un instant précis du dialogue.

### L'évolution future du système ARA

La raison pour laquelle l'utilisateur de la première version du système est obligé d'épeler les noms au moyen d'une table d'épellation réside principalement dans la nature du dispositif de reconnaissance des mots. En effet, seule cette méthode permet à l'utilisateur d'introduire une succession de lettres que le système ARA est en mesure de reconnaître avec une fiabilité suffisante.

Il serait, bien entendu, plus commode de pouvoir épeler normalement les noms (sans table d'épellation). Toutefois, vu que les lettres prononcées au téléphone ne peuvent être reconnues que de manière peu fiable, des informations supplémentaires sont indispensables pour améliorer la reconnaissance des caractères. Il est possible, en utilisant la matrice de confusion du dispositif de reconnaissance, de déterminer la meilleure succession de caractères correspondant à un nom possible. Cela exige, il est vrai, une liste des localités ou des listes comportant les noms des usagers ou des rues selon la localité, etc. Avec l'interface ETV actuellement disponible, cela n'est toutefois pas réalisable.

Il est possible de rendre l'utilisation du système ARA nettement plus agréable avec un dispositif de reconnaissance des mots indépendants d'un vocabulaire,

étant donné que ce dispositif peut reconnaître, de par son principe, des mots prononcés normalement et provenant d'un vocabulaire d'une dimension quelconque. Pour cela, il ne suffit pas de disposer des listes évoquées plus haut, mais encore d'un système connaissant la prononciation usuelle de chaque nom propre.

En réalisant ce pas au niveau du dispositif de reconnaissance de la parole et en adaptant le dialogue de manière optimale, il devrait être possible qu'un large public accepte le système ARA et qu'il l'utilise souvent. A la question de savoir pourquoi la résolution directe de ces divers problèmes n'a pas été incluse dès le début dans les objectifs du projet proprement dit, on peut répondre par ce qui suit: d'une part, l'étape, permettant de passer du point de départ à cet objectif, serait trop grande, c'est-à-dire, le problème comporterait trop d'inconnues pour pouvoir être valablement évalué. D'autre part, l'épellation avec ou sans table prescrite sera pratiquée de toute façon dans un système futur, comme on le fait très souvent et spontanément pour améliorer la compréhension lorsqu'on s'adresse à une opératrice d'un service desservi.

## Le synthétiseur de la parole SVOX

Le système SVOX (fig. 2) produit pour chaque texte allemand, phrase par phrase, un signal vocal. Il appartient donc à la catégorie du système texte-parole

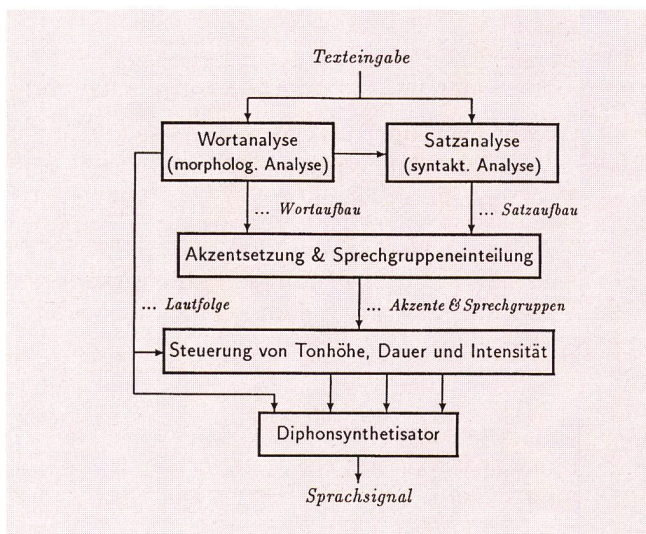


Fig. 2 Schéma bloc du système SVOX

Texteingabe – Entrée du texte  
 Wortanalyse (morpholog. Analyse) – Analyse des mots (analyse morphologique)  
 Satzanalyse (syntakt. Analyse) – Analyse des phrases (analyse syntaxique)  
 Wortaufbau – Structure du mot  
 Satzaufbau – Structure de la phrase  
 Akzentsetzung & Sprechgruppeneinteilung – Assignation des accents et groupement prosodique  
 Lautfolge – Succession des phonèmes  
 Akzente & Sprechgruppen – Accents et groupes prosodiques  
 Steuerung von Tonhöhe, Dauer und Intensität – Commande de la tonalité, de la durée et de l'intensité des sons  
 Diphonsynthetisator – Synthétiseur à diphonèmes  
 Sprachsignal – Signal vocal

(Text-to-Speech Systems). Il comprend deux parties principales, l'étape de transcription et l'étape phono-acoustique.

L'étape de transcription transpose l'entrée orthographique en une représentation phonologique (représentation abstraite du signal vocal à générer), qui comprend la succession des phonèmes, l'accentuation des mots et des phrases et les groupes prosodiques. A cet effet, on procède d'abord à une analyse morphologique des mots d'une phrase (détermination du genre et de la forme du mot) puis à sa transcription phonétique. Le résultat de l'analyse morphologique sert à déterminer la structure syntaxique de la phrase par l'emploi d'une grammaire. Cette structure sert à subdiviser l'énoncé en groupes prosodiques et à déterminer les accents relatifs des mots.

L'étape phono-acoustique détermine au vu de cette représentation abstraite la durée, la tonalité et l'intensité du signal vocal à synthétiser, ce que l'on appelle la commande prosodique. Enfin, le synthétiseur modifie les diphonèmes de manière qu'ils présentent la fréquence fondamentale, la durée et l'intensité désirées et puis finalement il les assemble pour produire le signal de sortie (synthèse par concaténation).

## L'interface ETV

L'interface avec l'annuaire électronique des abonnés a pour tâche principale de convertir les demandes de la commande centralisée en celles compatibles avec l'ETV et de transformer les résultats fournis par l'ETV sous une forme utilisable par le dialogue. Or, le système ETV actuel ne permet que partiellement une telle opération, parce que les informations en sortie ne sont pas structurées de manière qu'elles puissent être traitées sans une interprétation sémantique. Si un tel système d'interprétation est en principe réalisable, il ne pourra l'être qu'au prix d'un immense travail.

Pour la première version du système ARA, on n'a développé qu'un dispositif d'interprétation simple, dans l'hypothèse que les annuaires électroniques des abonnés des Télécom PTT seraient de toute façon mieux structurés dans un proche avenir.

## Le dispositif de reconnaissance des mots

La tâche et les difficultés des systèmes de reconnaissance de la parole sont expliquées dans les figures 3 et 4. La figure 3 représente un signal vocal d'une durée de deux secondes. Le locuteur prononce le nom de «Rosa». Avant et après le nom, on aperçoit des bruits perturbateurs typiques d'une liaison téléphonique. Sur la figure 4, on voit les spectres des segments temporels de 30 ms des 4 phonèmes correspondant au nom de «Rosa».

L'analyse des caractéristiques du signal vocal – et tout particulièrement la distribution spectrale de la puissance du signal ainsi que la dérivation temporelle de ces grandeurs – permettent une reconnaissance automatique de la parole. Pour traiter de telles caractéristiques vocales, on connaît des procédés recourant aux

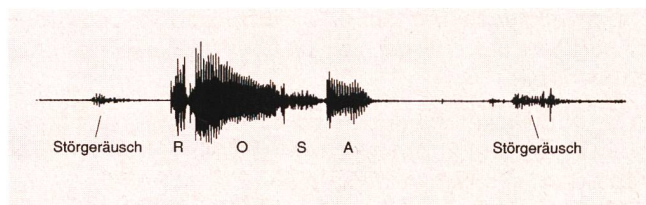


Fig. 3 Signal vocal du mot «Rosa»  
Störgeräusch – Bruit parasite

«modèles de Markov cachés» (Hidden Markov Models) [1] et les «réseaux neuromimétiques» [2]. Bien qu'il s'agisse ici de procédés classiques, une optimisation intensive de la structure et ses dimensions est nécessaire pour la réussite d'une application pratique.

A l'encontre de ce qui s'est fait lors de la phase expérimentale et de la première phase de test, dans lesquelles les débuts et les fins de mots utilisés étaient en règle générale marqués ou corrigés manuellement, dans la pratique, un dispositif de reconnaissance de la parole doit être en mesure de reconnaître les divers mots dans un signal continu. Une telle procédure est particulièrement difficile lors d'une liaison téléphonique, parce que les locuteurs se comportent parfois maladroitement, à savoir qu'ils produisent des bruits d'inspiration ou de claquement avant de prononcer les mots, auxquels s'ajoutent encore les bruits de la ligne et d'autres signaux perturbateurs.

Dans le cadre du projet ARA, divers travaux relatifs à la reconnaissance de la parole ont été réalisés par Ascom Tech AG, à savoir

- au laboratoire d'Hombrechtikon: l'amélioration d'un logiciel de reconnaissance de la parole provenant d'un avant-projet et la construction d'un dispositif de reconnaissance des mots réalisé avec un PC contenant une carte de traitement du signal (DSP) et une interface RNIS
- au laboratoire de Soleure: la reconnaissance de diverses lettres prononcées isolément au moyen de réseaux neuromimétiques.

#### Le logiciel de reconnaissance de la parole

Le point de départ des travaux portant sur le dévelop-

pement d'un logiciel de reconnaissance était un programme qui, compte tenu de débuts et de fins de mots prédéterminés, était capable de reconnaître les chiffres de 0 à 9 avec une fiabilité de 95 %. Lors des premiers travaux, il fut possible d'augmenter la fiabilité de la reconnaissance à 99 %. Les expériences qui s'ensuivirent avec les 26 mots de la table d'épellation ont confirmé le chiffre de 99 %. Au cours d'une deuxième étape, on a tenté de reconnaître dans un signal ininterrompu la succession de certains mots, c'est-à-dire sans que le début et la fin des mots soient marqués manuellement. Il fut ainsi possible d'atteindre finalement une fiabilité de la reconnaissance de 96 %.

#### La configuration matérielle utilisée pour la reconnaissance de la parole

L'emploi pratique souhaité de la reconnaissance des mots dans le système ARA a fait porter le choix sur une solution PC en raison des composants de système actuellement disponibles. Grâce à une carte RNIS en tant qu'interface avec SwissNet 2, on peut diriger des signaux vocaux continus via le bus PC sur la carte de traitement numérique des signaux avec une puissance de calcul de 30 MFLOPS (millions d'opérations de calculs en virgule flottante par seconde, Mega Floating-Point Operations per Second) pour effectuer la reconnaissance de la parole. On a en outre réalisé sur le PC la commande des cartes RNIS et DSP ainsi que le dialogue avec la commande d'application centralisée en tant que processus parallèles.

#### La reconnaissance de la parole par réseaux neuromimétiques de lettres prononcées isolément

Pour pouvoir procéder à ces travaux, il a fallu d'abord enregistrer 200 voix de locuteurs et locutrices et marquer les débuts et les fins des mots. On a ensuite procédé à des expériences avec des réseaux neuronaux temporisés (Time Delay Neural Networks, TDNN), qui sont spécialement adaptés à différentes catégories de lettres pouvant être confondues avec une grande vraisemblance, et qui peuvent ensuite être rassemblées en un réseau global. En prenant comme exemple l'ensemble de lettres «f, l, m, n, r, s», on est parvenu à un taux de reconnaissance de 86 %.

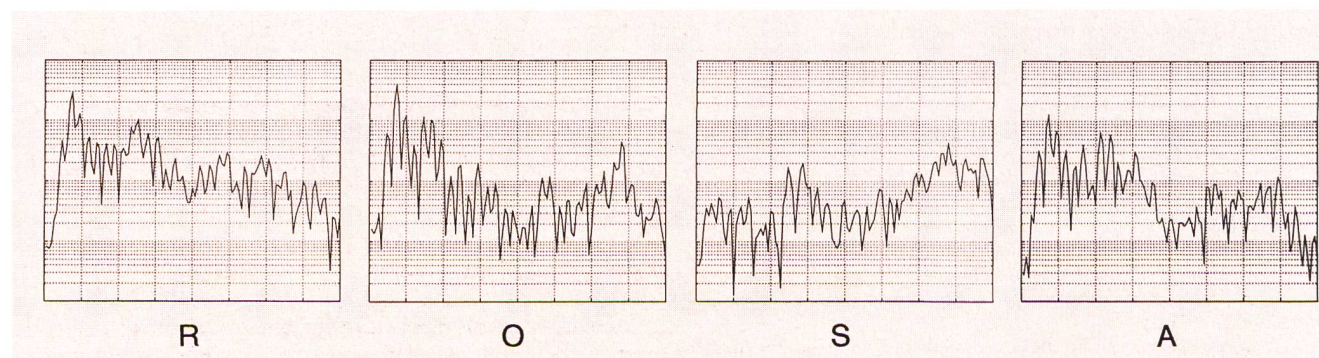


Fig. 4 Spectre des fréquences des phonèmes R, O, S et A

## L'utilité du projet ARA pour les Télécom PTT

Comme nous l'avons dit plus haut, divers systèmes de reconnaissance de la parole sont opérationnels à l'étranger, bien qu'il s'agisse d'applications beaucoup plus simples que celles relatives aux tâches que doit assumer un service de renseignements fournissant les numéros d'abonnés. Il ne s'agit pas ici de débattre des raisons pour lesquelles de tels systèmes n'ont pas encore été mis en service en Suisse pour l'instant.

Dans l'optique du projet ARA, il est toutefois important d'attirer l'attention du lecteur sur le fait qu'il est bien entendu facile d'intégrer des services simples dans le système ARA déjà en place. Tel est le cas pour ARA-V1, capable de fournir des services d'information tels que les prévisions météorologiques, l'horloge parlante et le service de réveil automatique. Des difficultés notables ne sont en fait prévisibles que pour des applications dans lesquelles le synthétiseur de parole doit convertir en signaux vocaux des textes

contenant des noms propres ou des noms de langues étrangères. Tel est certainement le cas pour les nouvelles et les communiqués sportifs.

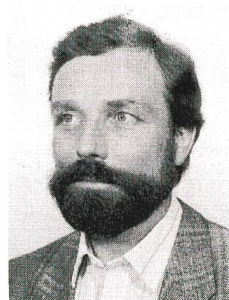
### Adresses des auteurs

Beat Pfister  
Institut für Technische Informatik  
und Kommunikationsnetze  
ETH Zentrum  
CH-8092 Zurich

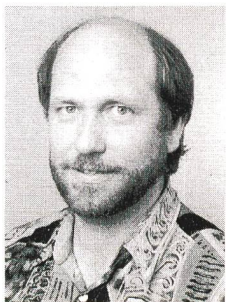
Arthur Schaub  
Ascom Tech AG, H 32 412  
CH-8634 Hombrechtikon

### Bibliographie

- [1] *Rabiner L.* A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. of the IEEE, Febr. 1989, S. 257-286.
- [2] *Weibel A. and Hampshire J.* Building Blocks for Speech. Byte, Aug. 1989. S. 235-242.



**Beat Pfister** est né en 1950 à Dagmersellen dans le canton de Lucerne. Après sa maturité scientifique à Lucerne, il obtient son diplôme d'ingénieur électricien à l'EPF de Zurich. Au début de 1976, il devient membre du nouveau groupe «Traitement de la parole» de l'institut de physique technique où il s'emploie surtout à réaliser des processeurs de forte puissance de calcul pour le traitement de signaux. Depuis 1981, il dirige le groupe «Traitement de la parole» rattaché à l'institut d'informatique technique et des réseaux de communication. Au cours des dernières années, ce groupe de huit collaborateurs s'est surtout occupé de la synthèse et de la reconnaissance en langue allemande et de leurs applications dans le domaine des systèmes de renseignements automatiques.



**Arthur Schaub**, né en 1950, a étudié la physique expérimentale à l'Université de Berne. A côté de la simulation numérique de circuits analogiques pour divers projets industriels, il se spécialise dans le traitement classique, adaptatif et multidébits des signaux numériques. Au printemps 1988, il entre chez Ascom où il dirige un laboratoire s'occupant du traitement numérique de la parole. Les activités principales consistent à coder la parole en vue de la mémoriser efficacement et de la transmettre sous forme numérique, de même qu'à reconnaître des messages prononcés par des locuteurs entraînés ou non ainsi qu'à conditionner des signaux dans le domaine de l'acoustique des appareils de correction auditive. Il s'intéresse en outre à la statistique mathématique ainsi qu'à la théorie des tests et au codage permettant de reconnaître et de corriger les erreurs.